Identification and analysis of lipid metabolism-related genes in allergic rhinitis

Qilei Tao
Naval Medical University (Second Military Medical University)

Yajing Zhu
Naval Medical University (Second Military Medical University)

Tianyu Wang
Naval Medical University (Second Military Medical University)

Yue Deng
Naval Medical University (Second Military Medical University)

Huanhai Liu
Naval Medical University (Second Military Medical University)

Jian Wu (jianwu2008@126.com)
Naval Medical University (Second Military Medical University)

Research Article

Keywords: Allergic rhinitis, Lipid metabolism, Biomarkers, Diagnostic model, Immune infiltration

Posted Date: January 12th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2461941/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Lipids in Health and Disease on July 21st, 2023. See the published version at https://doi.org/10.1186/s12944-023-01825-z.
Abstract

**Background:** Studies have shown that the lipid metabolism mediator leukotriene is associated with the pathogenesis of allergic rhinitis (AR). The aim of this study was to identify key lipid metabolism-related genes (LMRGs) related to the diagnosis and treatment of AR.

**Material and methods:** AR-related expression datasets (GSE75011, GSE46171) were downloaded through the Gene Expression Omnibus (GEO) database. First, weighted gene coexpression network analysis (WGCNA) was used to get AR-related genes (ARRGs). Next, between control and AR groups in GSE75011, differentially expressed genes (DEGs) were screened, and DEGs were intersected with LMRGs to obtain lipid metabolism-related differentially expressed genes (LMR DEGs). Protein–protein interaction (PPI) networks were constructed for these LMR DEGs. Hub genes were then identified through stress, radiality, closeness and edge percolated component (EPC) analysis and intersected with the ARRGs to obtain candidate genes. Biomarkers with diagnostic value were screened via receiver operating characteristic (ROC) curves. Differential immune cells screened between control and AR groups were then assessed for correlation with the diagnostic genes, and clinical correlation analysis and enrichment analysis were performed. Finally, reverse transcription-polymerase chain reaction (RT-qPCR) was made on blood samples from control and AR patients to validate these identified diagnostic genes.

**Results:** 73 LMR DEGs were obtained, which were involved in biological processes such as metabolism of lipids and lipid biosynthetic processes. Sixty-six ARRGs and 22 hub genes were intersected to obtain four candidate genes. Three diagnostic genes (LPCAT1, SGPP1, SMARCD3) with diagnostic value were screened according to the AUC > 0.7, with markedly variant between control and AR groups. In addition, two immune cells, regulatory T cells (TReg) and T follicular helper cells (TFH), were marked variations between control and AR groups, and SMARCD3 was significantly associated with TFH. Moreover, SMARCD3 was relevant to immune-related pathways, and correlated significantly with clinical characteristics (age and sex). Finally, RT-qPCR results indicated that changes in the expression of LPCAT1 and SMARCD3 between control and AR groups were consistent with the GSE75011 and GSE46171.

**Conclusion:** LPCAT1, SGPP1 and SMARCD3 might be used as biomarkers for AR.

Introduction

Allergic rhinitis (AR) is an infectious inflammatory disease mediated by IgE that affects approximately 10%~20% of the global population [1]. Exposure to inhaled allergens in susceptible individuals is a frequent precipitating factor for AR. The most common clinical symptoms include paroxysmal sneezing, nasal obstruction, rhinorrhea, and nasal itching, sometimes in association with conjunctivitis, such as eye itching and tearing. Persistent severe rhinitis may predispose patients toward asthma [2]. Moreover, AR patients often have decreased learning and work efficiency, impaired sleep and quality of life, and even psychological disorders such as depression, leading to a huge economic burden
on society. For treatment, nasal corticosteroids, antihistamines, and leukotriene receptor antagonists are currently the most recommended drugs [3]. However, their long-term drug use leads to a range of side effects, including epistaxis and drowsiness. Furthermore, the sustained poor efficacy of available drugs causes recurrent illnesses. Thus, it is crucial to find effective therapeutic targets for AR treatment.

Lipids are composed of fats and lipoids and play an important role in different organelles as a second messenger for intracellular signaling [4]. Lipid metabolism refers to the digestion, synthesis, and disassembly of lipids, with the help of various enzymes related to the processing of substances necessary for the body to ensure normal physiologic function. Previous studies have shown that lipid metabolism-related genes (LMRGs) are associated with several systemic diseases. For instance, Li et al. found that LMRGs in circulation have good predictive value for early diagnosis of intervertebral disc degeneration (IDD) [5]. LMRGs are also involved in lung cancer development and might serve as biomarkers for lung cancer [6], and the lipid compound prostaglandins is an effective therapeutic target in allergic airway diseases [7]. Leukotriene is a well-recognized lipid inflammatory mediator in allergic diseases, and leukotriene receptor antagonists are one of the major medications for AR [8]. In addition, AR patients have a high level of apolipoprotein in nasal mucus, which may be involved in lipid metabolism and have immunomodulatory properties [9]. Nevertheless, the relationship between LMRGs and AR has remained unclear.

In this study, AR-related public datasets and comprehensive bioinformatics methods were used to identify LMRGs with diagnostic value for AR, providing a potential treatment of AR patients.

**Materials And Methods**

**Data extraction**

AR-related datasets (GSE75011 and GSE46171) were downloaded through GEO database. The GSE75011 was the training set, containing 15 control and 25 AR blood samples. GSE46171 dataset containing 3 control and 6 AR samples of nasal mucous was used as an external validation set. A total of 750 LMRGs were gained through Reactome and the Kyoto Encyclopedia of Genes and Genomes (KEGG) repository [5].

**Identification Of Ar-related Genes (ARRGs)**

To gain ARRGs in GSE75011, weighted gene coexpression network analysis (WGCNA) was performed. First, the samples were clustered to remove outliers. Thereafter, the determination of soft threshold (β) was performed. Modules were segmented via dynamic tree cutting based on optimal β. Correlations were analyzed between modules and AR. The genes of the highest relevance module with |gene significance (GS)| > 0.3, |module membership (MM)| > 0.6, and \( P < 0.05 \) were defined as ARRGs [10].
Screening And Functional Analysis Of Lipid Metabolism-related Differentially Expressed Genes (Lmr Degs)

First, sample normalization on the GSE75011 dataset was performed via “limma” R package (version 3.48.3). The mRNA expression levels between control and AR groups in the GSE75011 dataset were contrasted via the “limma” R package (version 3.48.3) \( (P < 0.05) \) \cite{11}. DEGs and LMRGs were taken to intersect to get LMR DEGs. Subsequently, enrichment analysis for LMR DEGs via Metascape database \( (P < 0.05) \) \cite{12}. In addition, the online database WebGestalt was used to study the isease Ontology (DO) function of LMR DEGs.

Creation of protein–protein interaction (PPI) networks of LMR DEGs and screening of hub genes

PPIs of LMR DEGs were created via Search Tool for the Retrieval of Interacting Genes (STRING). Subsequently, Cytoscape was utilized to visualize PPIs, and hub genes were obtained by intersecting the top 30 genes calculated by stress, radiality, closeness, and edge percolated component (EPC).

Screening Of Diagnostic Genes

First, candidate genes were obtained by intersecting hub genes with ARRGs, and enrichment analysis were applied on them. Second, receiver operating characteristic (ROC) curves of the candidate genes were mapped via “pROC” R package (version 1.18.0) in the GSE75011 and GSE46171\cite{13}. Candidate genes with area under the curve (AUC) \( \geq 0.7 \) were regarded diagnostic genes. A nomogram was constructed with hub genes, and a calibration curve of the nomogram was drawn to verify its validity. The diagnostic worth of age, sex, and time point was assessed in GSE46171 via ROC curves.

Immune Analysis

The single-set gene set enrichment analysis (ssGSEA) algorithm was utilized to assess infiltrating richness of immune cells between AR and normal groups in the training set. Differences of the control and AR groups were compared by the Wilcoxon test. In addition, relevance was analyzed via Spearman algorithm between diagnostic genes and differential immune cells.

Analysis Of Clinical Correlation

Relevance between diagnostic genes and clinical characteristics (age, sex, time point) was analyzed using Pearson in the “corrplot” R package \cite{14}.

Gene Set Enrichment Analysis (Gsea) Of Diagnostic Genes
On the basis of the median value of the diagnostic genes expression, the samples of GSE75011 were grouped into high and low expression groups. All genes in two expression groups were performed GSEA with \(|\text{normalized enrichment score (NES)}| > 1\), nominal (NOM) \(P\) value \(< 0.05\), and \(q < 0.25\) [15].

**Patients And Tissue Preparation**

Ten AR patients and ten patients without AR or significant underlying disease were selected from people visiting to Shanghai Changzheng Hospital. There were no marked variation in sex and age between the groups (Table 1). Blood samples were acquired from these patients with informed consent and carried out reverse transcription-polymerase chain reaction (RT–qPCR). This Medical Ethics Committee of Shanghai Changzheng Hospital endorsed this study.

<table>
<thead>
<tr>
<th>Group</th>
<th>Sex</th>
<th>Age (mean ± STD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Control</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

STD, standard deviation

**Rt–qpcr**

Firstly, we conducted the total RNA extraction utilizing TRIzol (Ambion, Austin, USA). Then, reverse transcription of total RNA to cDNA was made via First-strand-cDNA-synthesis-kit (Servicebio, Wuhan, China). RT–qPCR was made utilizing the 2xUniversal Blue SYBR Green qPCR Master Mix (Servicebio, Wuhan, China). Specific experimental steps were carried out on the basis of instructions. The primer sequences were showcased in Additional file 1. Internal reference gene was GAPDH. The \(2^{-\Delta\Delta C_{t}}\) method was utilized to calculate the expression of diagnostic genes [16]. Levels of expression of diagnostic genes between the control and AR groups were compared by the T test.

**Statistical analysis**

Statistical analysis was carried out through GraphPad Prism 5 and R software (version 4.2.0). \(P < 0.05\) represented a significant difference. Differences between groups were analyzed via the Wilcoxon test.

**Results**

**Acquisition of ARRGs**
To identify ARRGs, WGCNA was performed with the GSE75011 dataset. Sample clustering analysis showed no outliers in the dataset (Fig. 1A). The β was 4 (Fig. 1B), and each gene module contained a minimum of 100 genes. Three modules were eventually identified, each with a unique color (Fig. 1C-D). The blue module correlated markedly with AR (cor = -0.35, P = 0.03) (Fig. 1D). Finally, 66 ARRGs were gained and utilized for further analysis (Fig. 1E).

**Acquisition And Functional Enrichment Of Lmr Degs**

The 25 samples were standardized for the GSE75011 dataset and are presented as box plots in Fig. 2A-B. The volcano plot and heatmap show 1621 DEGs between the AR and control groups, including 810 upregulated and 811 downregulated genes (Fig. 2C-D). A total of 73 LMR DEGs (Additional file 2) were obtained by Venn analysis with LMRGs (810 genes) and DEGs (1621 genes), with a significant difference detected based on a heatmap (Fig. 2E-F). Enrichment analysis of the 73 LMR DEGs by Metascape showed a total of 334 functional pathways (Fig. 2G-H) to be related to the LMR DEGs, such as metabolism of lipids, lipid biosynthetic process, and sterol regulatory element-binding protein (SREBP) signaling. DO enrichment results showed that the LMR DEGs are significantly associated with 10 diseases, namely, xanthomatosis, increased serum pyruvate, decreased high-density lipoprotein, hypoalphalipoproteinemia, myoglobinuria, insulin-resistant diabetes, neonatal death, thin skin, myalgia and cardiomegaly (Fig. 2I).

**Acquisition Of Hub Genes**

The PPI network was created for LMR DEGs. As illustrated in Fig. 3A-B, SREBF1 interacts with multiple proteins, such as LPIN1, GPAM, and MED1. To identify the most important genes, the 22 genes common to the 4 algorithms were used as hub genes (Fig. 3C), and a PPI network of hub genes was created (Fig. 3D). The results showed that GPAM interacts with 7 genes, namely, PPARG, NFYA, SREBF1, ACSL3, LPIN1, HMGCS1 and AACS.

**Acquisition Of Diagnostic Genes**

Four candidate genes associated with diagnosis of AR were obtained by 66 ARRGs with 22 hub genes taking intersections: LPCAT1, SREBF1, SMARCD3, and SGPP1 (Fig. 4A). The four candidate genes were involved in 133 GO items, including 114 GO BP, 9 GO CC and 10 GO MF, such as retina development in camera-type eye, npBAF complex, and transcription coregulator binding (Fig. 4B).

The diagnostic value of four candidate genes was assessed via ROC curve in GSE75011 and GSE46171. The AUC values for the three genes (LPCAT1, SMARCD3, and SGPP1) were greater than 0.7 in both datasets, suggesting that the three genes have diagnostic value for AR (Fig. 4C-D). The AUCs for age, sex and time point were 0.4492, 0.4839, and 0.7166, respectively, in GSE46171, revealing that sex might be a diagnostic factor for AR (Fig. 4E).
Finally, the nomograms were created containing the three diagnostic genes in GSE75011 and GSE46171 (Fig. 5A-B), and the AUC values in both datasets were above 0.6 (Fig. 5C-D). The results demonstrated that the nomogram has good prediction ability for AR.

**Immuno-infiltration Analysis In Ar And Control Groups**

Analysis of the percentage of immune cells by ssGESA in all samples showed the highest for T cells (Fig. 6A). Differences in infiltrating immune cells between the AR and control groups were illustrated by a violin plot (Fig. 6B). The results suggested that infiltration of regulatory T cells (TRegs) and T follicular helper cells (TFHs) was markedly lower in the AR group. There was significant relevance between SMARCD3 and TFHs. However, neither LPCAT1 nor SGPP1 correlated with differential immune cells (TRegs and TFHs); therefore, SMARCD3 was selected for further analysis (Fig. 6C-E).

**Correlation Analysis Of Clinical Features, Enrichment Analysis And Infiltration Analysis Of Smarcd3**

Pearson correlation analysis demonstrated that SMARCD3 was significantly associated with clinical characteristics (age and sex) (Fig. 7A-C). Then, GSEA for SMARCD3 was performed, revealing 256 GO enrichment (Additional file 3) and 33 KEGG (Additional file 4) pathways (Fig. 7D-E). Overall, SMARCD3 was involved in immune-related pathways, for instance, the B-cell receptor signaling pathway and T-cell receptor signaling pathway. Four immune cells displayed marked variations between the high and low expression groups, namely, macrophages, T helper cells, Tcm, and TFH cells, reflecting the strong relevance between SMARCD3 and the immune microenvironment (Fig. 7F).

**Mrna Levels Of Diagnostic Genes**

The significant differences in expression of SGPP1, LPCAT1 and SMARCD3 between control and AR in GSE75011 and GSE46171 were clearly observed via visualized data (Fig. 8A-B). Moreover, the changes of the three genes expression were consistent in blood and nasal mucosal tissues, suggesting that these three genes are of high diagnostic value.

To verify diagnostic gene expression, we collected blood samples to assess mRNA expression levels of three prognostic genes via RT-qPCR. The expression trends of LPCAT1 and SMARCD3 were consistent with public databases, and the expression was lower in AR group (Fig. 9A-B). However, SGPP1 exhibited the opposite trend compared to the results of public database, possibly due to different experimental designs or analysis methods (Fig. 9C).

**Discussion**
AR is an airway allergic disease with a high incidence, affecting billions of people in the world. Nevertheless, the effect of current therapies for AR is unsatisfactory due to its complex pathogenesis. LMRGs are involved in the maintenance of systemic physiology and play an important role in diverse diseases, especially in malignant tumors. Moreover, lipid-related inflammatory mediators such as prostaglandins and leukotrienes have been implicated in AR pathogenesis. To our knowledge, this is the first study to identify and analyze LMRGs in AR.

In this study, three key LMRGs most associated with AR, i.e., LPCAT1, SGPP1, and SMARCD3, were identified, all of which are protein-coding genes. As one of the lysophosphatidylcholine acyltransferase (LPCAT) family, the LPCAT1 protein is an enzyme essential for phosphatidylcholine metabolism and regulation of phosphatidylcholine composition [17]. LPCAT1 is also used in the prognosis of multiple tumors, such as breast cancer, colorectal cancer, and hepatocellular carcinoma [18–20]. Little is known about LPCAT1 in allergic diseases. One study reported that LPCAT1 downregulates eosinophilic inflammation in asthmatic mice [21]. In the current study, LPCAT1 was significantly lower in AR blood samples, consistent with published results, suggesting that it may be essential for AR pathogenesis.

SGPP1 can catalyze degradation of S1P, who can regulate diverse biological processes, as a bioactive sphingolipid metabolite [22]. SGPP1 is considered to be closely related to several tumors, especially regarding chemoresistance and radioresistance [23, 24]. There are currently no reports about the function of SGPP1 in allergic diseases, and the results in the current study are the first to show significant downregulation of SGPP1 in both blood and nasal mucosa samples in AR patients; conversely, RT-qPCR using blood samples showed the opposite result, possibly due to different experimental designs or analysis methods. Thus, the effect of SGPP1 in AR is still unclear. SMARCD3 is a chromatin-remodeling factor and a member of the SWI/SNF family, which present helicase and ATPase activities and are crucial in the transcription process of certain genes. Its related pathways include the circadian clock and transcriptional activation of mitochondrial biogenesis [25]. SMARCD3 was found to be downregulated in AR patients in this study, but how it participates in disease processes remains to be explored.

Immuno-infiltration analysis refers to studying the composition and quantification of immune cells in diseases. In this study, T follicular helper cells (TFHs) were extremely significantly reduced in the AR group. TFHs are CD4+ T cells that specialize in helping B cells and are involved in a wide range of diseases. An increasing number of theories have concluded that the antigen-related IgE response depends on more TFHs than Th2 cells [26, 27]. There are few reports about TFH and SMARCD3. A microarray model system identified that the SMARCD3 gene is upregulated in T-cell acute lymphoblastic leukemia [28]. In this study, only SMARCD3 correlated with differential immune cells (TRegs and TFHs), and TFHs and SMARCD3 were downregulated simultaneously in AR patients. Hence, it is hypothesized that SMARCD3 participates in the differentiation of T cells.

GSEA was performed to further investigate the role of SMARCD3 in AR, the results of which showed significant enrichment in the adipocytokine signaling pathway, B-cell receptor signaling pathway, and chemokine signaling pathway, among others. The adipocytokine signaling pathway refers to a series of cascade events via autocrine or paracrine adipocytokines, such as leptin and adiponectin, by adipocytes
in the body [29, 30]. This pathway is not only crucial for obesity, insulin resistance, and type II diabetes mellitus but also plays an important role in inflammation and allergic diseases. Dysregulation of pulmonary adipocytokine/insulin signaling caused by early-onset obesity has been proven to induce asthma-like disease in mice [31]. The leptin/osteopontin axis promotes Th2 inflammation and Th17 responses in AR through the NF-κB, MAPK, JNK pathway and β3 integrin [32, 33]. Signaling through the B-cell receptor (BCR) is crucial for antigen recognition and subsequent biological effects, including B-cell activation, proliferation, and differentiation, which ensure host defense [34]. One study demonstrated that the BCR signaling pathway was significantly enriched among differentially expressed vesicle miRNAs in AR patient nasal mucus, consistent with the findings in the current study and further elucidating the importance of the BCR signaling pathway in AR development [35]. Chemokines are small molecule-scale cytokines that recruit leukocyte subsets under steady-state and pathological conditions; signaling pathways are activated by their binding to receptors on the cell surface and are involved in chronic inflammatory and autoimmune diseases. Multiple studies have shown that knockdown of the chemokine receptor CCR3 reduces eosinophilic inflammation and the Th2 immune response in AR [36–38]. In summary, our findings are in accordance with all of the above studies.

**Comparisons With Other Studies And Contribution Of The Current Work To Existing Knowledge**

To the best of our knowledge, exploration of AR based on GSE75011 and GSE46171 has mainly targeted key genes differentially expressed between AR and control samples [39–42]. In the current study, the biological significance of lipid metabolism in AR was first systematically explored at the genetic level through these datasets. Moreover, correlation between SMARCD3 expression and immune cell infiltration was investigated to elucidate the underlying role of immune-related treatment targeting the SMARCD3 gene in exploration of AR development.

**Study Strengths And Limitations**

Three key LMRGs with high diagnostic values for AR were identified and analyzed for the first time based on bioinformatics analysis of AR-related expression datasets. However, the limitations of this study cannot be ignored. First, small sample sizes and small datasets of AR may have introduced bias. Second, the mechanisms of these genes in AR development have not been clearly elucidated. Further research is needed for the possibility of clinical use in the future.

**Conclusions**

In summary, this is the first bioinformatics analysis of LMRGs in AR, and three key genes (LPCAT1, SGPP1 and SMARCD3) with high diagnostic value for AR were identified. A highly accurate nomogram was constructed to validate the clinical applicability of the gene-based diagnostic model. In addition, two of these genes were confirmed by clinical validation and are considered potential treatment targets. In
particular, the correlation of SMARCD3 expression and immune cell infiltration was helpful to reveal future research directions of immune-related treatment targeting the SMARCD3 gene in AR.

**Abbreviations**

AR: allergic rhinitis; LMRGs: lipid metabolism-related genes; GEO: Gene Expression Omnibus; WGCNA: weighted gene coexpression network analysis; ARRGs: AR-related genes; DEGs: differentially expressed genes; PPI: protein–protein interaction; EPC: edge percolated component; ROC: receiver operating characteristic; RT–qPCR: reverse transcription-polymerase chain reaction; IDD: intervertebral disc degeneration; KEGG: Kyoto Encyclopedia of Genes and Genomes; TOM: topological overlap matrix; GS: gene significance; MM: module membership; GO: Gene Ontology; DO: Disease Ontology; STRING: Search Tool for the Retrieval of Interacting Genes; AUC: area under the curve; ssGSEA: single-set gene set enrichment analysis; GSEA: gene set enrichment analysis; NES: normalized enrichment score; NOM: nominal; GAPDH: glyceraldehyde-3-phosphate dehydrogenase; SREBP: sterol regulatory element-binding protein; BP: biological process; CC: cellular component; MF: molecular function; TReg: regulatory T cell; TFH: T follicular helper cell; Tcm: central memory T cell; BCR: B-cell receptor

**Declarations**

**Acknowledgments**

We thank the GEO database for providing high-quality data, as well as authors for making their valuable research results publicly available.

**Author contributions**

QT and JW designed and conducted the study. YZ and TW collected and analyzed the data. QT and YZ collected the blood samples and carried out the RT–qPCR tests. QT and YD drafted the manuscript. HL and JW critically reviewed this manuscript. All authors contributed to and approved the final manuscript.

**Funding**

Not applicable.

**Availability of data and materials**

GSE75011 and GSE46171 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/).

**Ethics approval and consent to participate**

Informed consent was obtained from all participants, and all experimental protocols were approved by the Medical Ethics Committee of Shanghai Changzheng Hospital.

**Consent for publication**
Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1Department of Otolaryngology, Changzheng Hospital, Naval Medical University (Second Military Medical University), Shanghai 200003, China.

**References**


**Figures**

**Figure 1**

**Gene coexpression network of allergic rhinitis.** (A) Sample clustering analysis of the GSE75011 dataset. (B) Analysis of network topology for various soft-thresholding powers showed that the soft threshold was 4. (C) Clustering dendrograms and modules identified by WGCNA, where the minimum gene number was 100. (D) Module-trait relationships demonstrated that the blue module correlated markedly with AR (cor = -0.35, \( P = 0.03 \)) in three modules, and each square contains the corresponding correlation and \( P \) value. (E) Correlation scatterplot of 66 ARRGs with |\( \text{GS} \)| > 0.3, |\( \text{MM} \)| > 0.6, and \( P < 0.05 \) in the blue module.
Figure 2

Acquisition and analysis of LMR DEGs. (A-B) The 25 samples of the GSE75011 dataset were standardized and presented as box plots. (C) Volcano plot of 1621 DEGs between the AR and control groups ($P < 0.05$). (D) Heatmap of 1621 DEGs between the AR and control groups ($P < 0.05$). (E) Venn diagram of 73 LMR DEGs by overlapping LMRGs and DEGs. (F) Heatmap of the 73 LMR DEGs. (G-H) Enrichment analysis of the 73 LMR DEGs in Metascape was performed, and the associated interaction
network is shown. (I) Disease Ontology (DO) function analysis of LMR DEGs using the WebGestalt database.

Figure 3

**Acquisition of hub genes.** (A) The protein–protein interaction (PPI) network for 73 LMR DEGs through STRING. (B) The degree of connectivity of each gene in the PPI network demonstrated that SREBF1
interacted with multiple proteins. (C) Venn diagram of hub genes common to four algorithms. (D) PPI network of hub genes.

Figure 4

Acquisition of diagnostic genes. (A) Venn diagram of four candidate genes by taking the intersections of ARRGs and hub genes. (B) Gene Ontology (GO) enrichment analysis of the four candidate genes. (C-D) AUC curves of the 3 genes for diagnostic prediction in the GSE75011 and GSE46171 datasets. (E) AUCs of age, sex, and time point in GSE46171.
Figure 5

(A-B) The nomogram drawn based on the three diagnostic genes for the diagnostic efficacy of AR in GSE75011 and GSE46171. (C-D) Calibration curves of the nomogram in GSE75011 and GSE46171.
Figure 6

**Immuno-infiltration analysis in AR.** (A) Analysis of the percentage of immune cells in all samples of GSE75011 using ssGESA. (B) Differences in immune cell infiltration between AR and control groups (Wilcoxon Test). **P < 0.01. Lollipop chart demonstrating the correlation between diagnostic genes and immune cells, including (C) LPCAT, (D) SGPP1, and (E) SMARCD3.
**Figure 7**

**Comprehensive analysis of SMARCD3 in AR.** (A-C) Pearson correlation analysis of SMARCD3 clinical characteristics (age, sex and time point). GSEA of SMARCD3 showed that 256 GO enrichment pathways (D) and 33 KEGG pathways (E) were enriched in AR. (F) Differences in immune cell infiltration between high and low SMARCD3 expression groups (Wilcoxon test), *$P < 0.05$; **$P < 0.01$.**
Figure 8

Expression levels of three diagnostic genes in online datasets. (A) GSE75011. (B) GSE46171. *P < 0.05; **P < 0.01.

Figure 9
Validation of SMARCD3 expression by reverse transcription-polymerase chain reaction (RT–qPCR). (A) SMARCD3. (B) LPCAT1. (C) SGPP1. *$P < 0.05$; ***$P < 0.001$

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Suppl.materials1.TableS1..xlsx
- Suppl.materials2.TableS2..xlsx
- Suppl.materials3.TableS3..xlsx
- Suppl.materials4.TableS4..xlsx
- Suppl.materials5.iThenticaterereport.pdf
- Suppl.materials6.SNASEditingCertificate.pdf